Tạp chí Khoa học Nông nghiệp Việt Nam 2025, 23(7): 899-909 https://doi.org/10.31817/tckhnnvn.2025.23.7.06
vie.vjas.vn

SO SÁNH MỘT SỐ PHƯƠNG PHÁP TRÍCH CHỌN ĐẶC TRƯNG VÀ PHÂN LỚP CHO DỮ LIỆU ÂM THANH ONG

Nguyễn Thị Huyền*, Nguyễn Văn Hoàng

Khoa Công nghệ thông tin, Học viện Nông nghiệp Việt Nam

**Tác giả liên hệ: nthuyen@vnua.edu.vn*

Ngày nhận bài: 25.11.2024 Ngày chấp nhận đăng: 18.06.2025

TÓM TẮT

Trong bài viết này, chúng tôi tiến hành thử nghiệm và đánh giá một số phương pháp trích chọn đặc trưng cho dữ liệu âm thanh ong. Các đặc trưng được trích chọn bằng các kỹ thuật MFCC, Chroma, Wavelet và các đặc trưng quan trọng được lựa chọn bằng phương pháp Rừng ngẫu nhiên, Extra trees, XGBoost từ dữ liệu ban đầu sau đó các đặc trưng này được cung cấp cho các thuật toán học máy như SVM, Rừng ngẫu nhiên, XGBoost để giải quyết bài toán nhận dạng tiếng ong. Kết quả thử nghiệm cho thấy với mô hình Rừng ngẫu nhiên và XGBoost sử dụng đặc trưng MFCC việc phân biệt tiếng ong là hoàn toàn khả thi với độ chính xác trên 99,9%.

Từ khóa: MFCC, Chroma, Wavelet, Rừng ngẫu nhiên, XGBoost, phân loại âm thanh ong.

Comparison of some Feature Extraction and Classification Methods for Bee Sound Data

ABSTRACT

In this study, we have tested and evaluated several feature extraction methods of bee sound data. The features were extracted by MFCC, Chroma and Wavelet techniques and important features were selected by Random Forest (RF), Extra trees, and XGBoost method from raw data and then provided to machine learning algorithms such as SVM, Random Forest, XGBoost to solve the bee recognition problem. The experiment results show that using Random Forest and XGBoost models with MFCC features, bee sound recognition is fully possible with the accuracy over 99.9%.

Keywords: MFCC, Chroma, Wavelet, Random Forest, XGBoost, Beehive sound recognition.

1. ĐẶT VẤN ĐỀ

Trong quá trình nuôi ong, theo dõi tổ ong để thu thập thông tin về trạng thái, hoạt động và tập tính của ong mật là một trong những công việc quan trọng nhất. Tuy nhiên, phương pháp thủ công thực hiện trong thực tế cần phải xâm lấn trực tiếp vào tổ ong nên gây ra sự căng thẳng, hoảng sợ cho ong, bên cạnh đó phương pháp này tốn nhiều thời gian và phụ thuộc chủ yếu vào kinh nghiệm, kiến thức của người nuôi ong. Chính các yếu tố này là động lực để các nhà nghiên cứu tìm ra các phương pháp dựa trên công nghệ hiện đại nhằm giảm bớt những nhược điểm của phương pháp truyền thống trong việc theo dõi đàn ong. Ý tưởng chung của cách tiếp cận mới này là sử dụng một hệ thống thiết bị điện tử gắn vào mỗi tổ ong để thu thập dữ liệu liên quan tới đàn ong như nhiệt độ, độ ẩm, trọng lượng buồng ong, âm thanh phát ra, số lượng ong bay ra - vào trong một khoảng thời gian. Dữ liệu thu thập được sau đó được phân tích và xử lý để thu được thông tin quan trọng cho phép người nuôi ong theo dõi tổ ong từ xa mà không làm gián đoạn vòng đời của các đàn ong mật (Du & cs., 2020).

Trong số các loại dữ liệu được thu thập từ tổ ong, âm thanh do đàn ong phát ra (tiếng ong) đóng vai trò quan trọng trong việc giám sát tổ ong tự động. Điều này là do tiếng vo ve của ong

mang thông tin về hành vi của đàn ong. Ví dụ, ong mật phát ra âm thanh cụ thể khi thiếu ong chúa hoặc chúng tiếp xúc với các tác nhân gây căng thẳng như ve ăn thịt và chất độc trong không khí (Bromenshenk & cs., 2009). Các phương pháp dựa trên công nghệ mới có thể giúp người nuôi ong giám sát tổ ong dựa trên âm thanh từ đó trích xuất thông tin quan trọng về hành vi của đàn mà không cần kiểm tra xâm lấn tổ ong. Nhiệm vụ cơ bản đầu tiên của bất kỳ công nghệ giám sát tổ ong dựa trên âm thanh nào là nhận ra âm thanh của ong và phân biệt chúng với âm thanh không phải của ong có thể thu được trong môi trường xung quanh tổ ong như âm thanh đô thị, mưa hoặc động vật khác như tiếng dế. Một khi hệ thống không phân biệt được âm thanh của ong với những âm thanh không phải của ong, tất cả các công việc liên quan đến phân tích dữ liệu cho các vấn đề cụ thể để thực hiện giám sát tổ ong tự động dựa trên âm thanh thu được từ tổ ong cũng sẽ thất bại.

Những năm gần đây, các phương pháp trích chọn đặc trưng âm thanh và kỹ thuật học máy được sử dụng khá nhiều trong nghiên cứu về dữ liệu âm thanh ong như: Terenzi & cs. (2019) đã nghiên cứu hiệu suất của phép biến đổi Wavelet (WT) như một kỹ thuật trích chọn đặc trưng để phân loại âm thanh của ong mật. Robles-Guerrero & cs. (2017) đã sử dụng mô hình hồi quy logistic (LR - Logistic Regression) và đặc trưng MFCC để phân biệt tiếng vo ve của ong mật và tiếng ồn. Kulyukin & cs. (2018) đã sử dụng kết hợp các đặct trưng MFFC, Chroma, Mel Spectrogram và Tonnetz với tổng cộng 193 đặc trưng làm đầu vào cho bốn thuật toán học máy để huấn luyện mô hình phân loại để phát hiện âm thanh của ong từ các mẫu âm thanh, bao gồm thuật toán LR, thuật toán K-Nearest Neighbor (KNN), thuật toán Rừng ngẫu nhiên (RF) và thuật toán SVM. Phương pháp Máy vectơ hỗ trợ (SVM) với nhân Gaussian được Nolasco & cs. (2019) sử dụng để phân loại dữ liệu từ các trạng thái khác nhau của tổ ong.

Trong nghiên cứu này, chúng tôi thực hiện thử nghiệm và đánh giá một số kỹ thuật trích xuất đặc trưng từ dữ liệu âm thanh ban đầu thu được từ các tổ ong. Các đặc trưng được trích chọn bằng các kỹ thuật MFCC, Chroma, Wavelet và các đặc trưng quan trọng được lựa chọn bằng phương pháp Rừng ngẫu nhiên, Extra trees, XGBoost từ dữ liệu thô. Sau đó những đặc trưng này sẽ được cung cấp cho các thuật toán học máy như SVM, Rừng ngẫu nhiên, XGBoost để giải quyết bài toán nhận dạng tiếng ong.

2. PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Vật liệu

2.1.1. Dữ liệu âm thanh ong

Chúng tôi sử dụng dữ liệu âm thanh ong được thu thập bởi (Kulyukin & cs., 2018) từ tháng 4/2017 đến tháng 9/2017 trong các tổ ong mật Ý ở hai trại nuôi ong cách nhau 17km. Trại nuôi ong thứ nhất ở môi trường đô thị, dữ liệu được thu nhận từ hai tổ cách nhau khoảng 20m: tổ 1.1 đặt bên cạnh một nhà để xe có máy phát điện; tổ 1.2 đặt gần một bãi xe lớn hơn, không có tiếng máy phát điện nhưng nhiều tiếng ồn như tiếng động cơ ô tô và tiếng còi xe. Trại nuôi ong thứ hai ở vùng nông thôn, dữ liệu được thu nhận từ hai tổ ong 2.1 và 2.2 cách nhau khoảng 15m. Mỗi tổ ong được đặt bên cạnh một tổ ong không được giám sát. Tổ 2.1 được đặt gần một bãi cỏ lớn có tiếng ồn từ máy cắt cỏ và hệ thống tưới nước tự động. Tổ 2.2 được đặt ở nơi yên tĩnh và hầu như không nghe thấy tiếng ồn xung quanh.

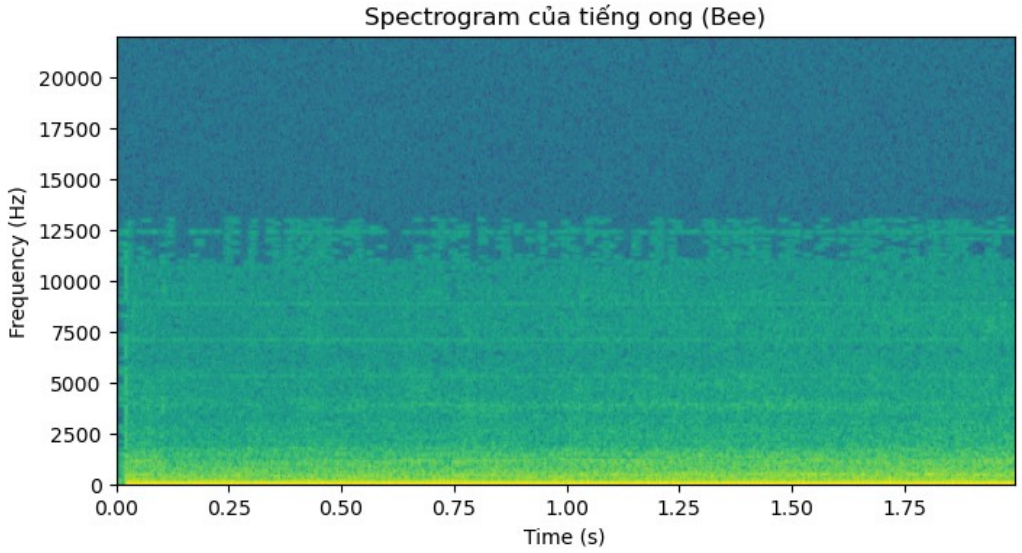
Để thu được dữ liệu âm thanh từ những tổ ong này Kulyukin & cs. (2018) đã thiết kế một hệ thống EBM đa cảm biến gồm nhiều thiết bị được kết nối với nhau: máy tính Raspberry Pi, máy ảnh thu nhỏ, bộ chia micrô, bốn micrô được kết nối với bộ chia, bảng điều khiển, tấm năng lượng mặt trời, cảm biến nhiệt độ, pin và đồng hồ phần cứng. Bốn micro có dải tần từ 15-20KHz được đặt cách cửa thùng 10cm và được gắn vào thành phía trong thùng ong. Hệ thống thực hiện thu nhận âm thanh 15 phút một lần, mỗi lần 30 giây. Mỗi tệp âm thanh có độ dài 30 giây với định dạng .wav sau đó được cắt thành từng đoạn có độ dài 2 giây với phần chồng chéo 1 giây. Mỗi mẫu âm thanh này sau đó được gán nhãn thuộc một trong ba lớp dữ liệu: tiếng ong vo ve (B), tiếng dế kêu (C) và

tiếng ồn xung quanh (N). Các hình 1-3 là hình ảnh quang phổ của tiếng ong vo ve (B), tiếng dế kêu (C) và tiếng ồn xung quanh (N), được tính từ ba mẫu âm thanh trong tập BUZZ1. Trong đó, thời gian được biểu thị dọc theo trục x, tần số dọc theo trục y và cường độ của tần số tại mỗi khung thời gian được biểu thị bằng màu sắc.

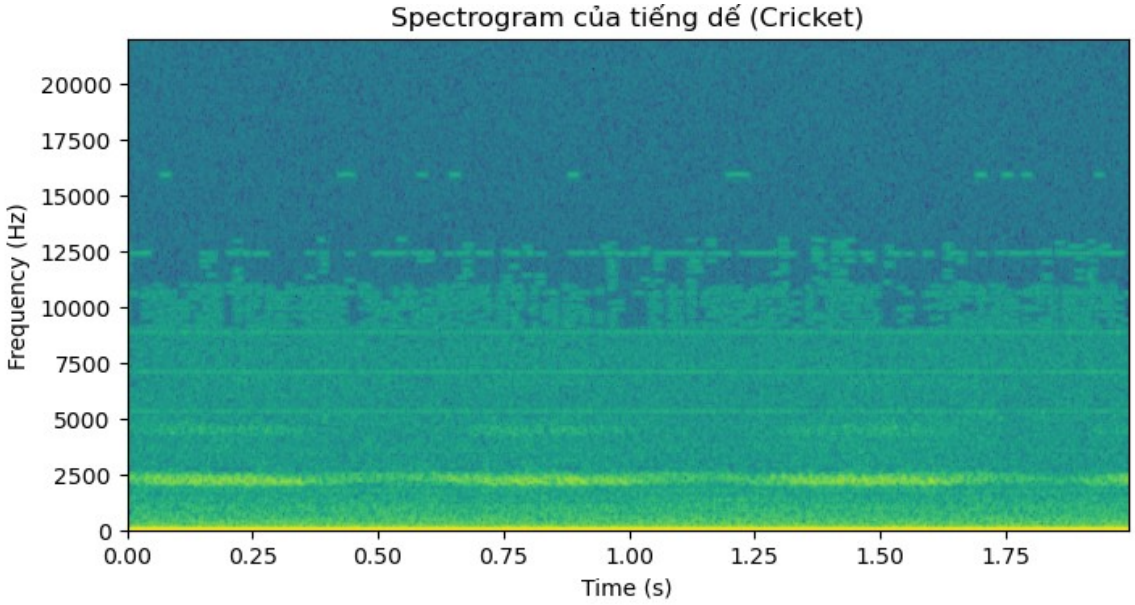
Dữ liệu thu nhận từ các tổ ong được đưa vào hai tập dữ liệu:

Tập BUZZ1 bao gồm tất cả 9.110 mẫu, được thu nhận từ các tổ ong 1.1 và 2.1 trong đó có: 3.000 mẫu B, 3.000 mẫu C và 3.110 mẫu N.

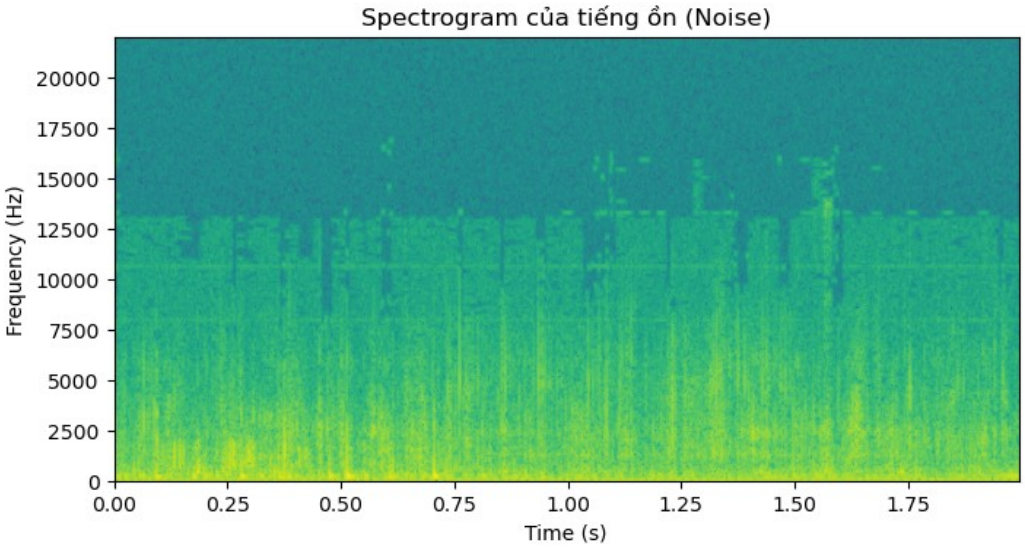
Tập BUZZ2 gồm tất cả 9.914 mẫu âm thanh, đã được chia thành hai tập: tập dữ liệu huấn luyện có 7.582 mẫu (chiếm 76,48%) được lấy từ tổ 1.1, trong đó bao gồm 2.402 mẫu B, 3.000 mẫu C và 2.180 mẫu N. Tập kiểm thử có 2.332 mẫu được lấy từ tổ 2.1, trong đó bao gồm 898 mẫu B, 500 mẫu C và 934 mẫu N. Như vậy, hai tập huấn luyện và kiểm thử hoàn toàn tách biệt nhau về tổ ong và vị trí đặt tổ.



Hình 1. Hình ảnh quang phổ của mẫu âm thanh ong vo ve



Hình 2. Hình ảnh quang phổ của mẫu âm thanh tiếng dế



Hình 3. Hình ảnh quang phổ của mẫu âm thanh tiếng ồn

2.1.2. Công cụ

Chúng tôi sử dụng máy tính Windows 64-bit, Intel(R) Core(TM) i7-10750H CPU @ 2.60GHz, 8Gb RAM, cài đặt các thuật toán bằng Python 3.8 (Python, 2021) và scikit-learn (Scikit-Learn, 2021).

2.2. Phương pháp nghiên cứu

2.2.1. Một số phương pháp trích chọn đặc trưng âm thanh

Để sử dụng được các kỹ thuật học máy tiêu chuẩn, yêu đầu tiên đặt ra là phải trích chọn đặc trưng để chuyển đổi tín hiệu thô thành vectơ đặc trưng. Vectơ đặc trưng là vectơ số đại diện cho một mẫu dữ liệu và được sử dụng trong nhiều tính toán hay phân loại.

a. Mel Frequency Cepstral Coefficient

Mel Frequency Cepstral Coefficient (MFCC) là một phương pháp trích chọn đặc trưng cho dữ liệu dạng âm thanh được giới thiệu bởi (Davis & Mermelstein, 1980). MFCC mô phỏng lại quá trình cảm nhận âm thanh của tai người dựa trên đặc điểm tai người nhận biết được những âm thanh có tần số thấp (< 1kHz) tốt hơn những âm thanh có tần số cao. Quá trình này được thực hiện thông qua các bước:

(1) Chuyển đổi tín hiệu tương tự sang kỹ thuật số và lọc tăng cường/nhấn mạnh trước: Âm thanh là một dạng tín hiệu liên tục, trong khi đó máy tính lại làm việc với các con số rời rạc. Vì vậy, ta cần chuyển từ dạng tín hiệu liên tục sang dạng rời rạc bằng cách lấy mẫu tại các khoảng thời gian cách đều nhau với một tần số lấy mẫu xác định.

(2) Phân đoạn: Quá trình này thực hiện phân đoạn các mẫu tín hiệu âm thanh thành các khung nhỏ theo thời gian với độ chồng chéo khoảng 50% giữa các khung liên tiếp nhau.

(3) Biến đổi Fourier nhanh: sử dụng biến đổi Fourier nhanh để biến đổi tín hiệu từ miền thời gian sang miền tần số (phổ biên độ).

(4) Các bộ lọc Mel: Để mô phỏng lại quá trình cảm nhận âm thanh của tai người, ở bước này ta xây dựng một thang Mel và một tập hợp (thường gồm từ 20 đến 40) các bộ lọc hình tam giác được sử dụng để tính tổng trọng số của các bộ lọc thành phần sao cho đầu ra của quá trình xấp xỉ với thang Mel.

(5) Biến đổi Cosine rời rạc: Đây là quá trình chuyển đổi phổ log Mel thành miền thời gian sử dụng biến đổi Cosin rời rạc (DCT). Kết quả của quá trình chuyển đổi này được gọi là Mel Frequency Cepstrum Coefficient.

b. Biến đổi Wavelet

Phép biến đổi wavelet liên tục:

Phép biến đổi Wavelet (WT) là phép biến đổi cung cấp biểu diễn tần số theo thời gian của

tín hiệu. Gọi f(x) là tín hiệu 1-D, phép biến đổi Wavelet liên tục là sự phân rã của f(x) thành một tập hợp hàm cơ sở $\Psi_{s,b}(x)$ được gọi là các Wavelet (Abdalla & Ali, 2010). Các Wavelet được tạo ra từ một Wavelet mẹ duy nhất Ψ(x) bằng cách thực hiện giãn nở và dịch.

$$w(a,b) = \int f(x)\Psi^*_{s,b}(s)dx \quad (1)$$

$$\Psi_{s,b}(x) = \frac{1}{\sqrt{s}}\Psi\left(\frac{x-b}{s}\right) \quad (2)$$

Trong đó: f(x) là tín hiệu cần phân tích, w(s, b) là hệ số biến đổi Wavelet liên tục của f(x), với s là tỉ lệ và b là dịch chuyển đặc trưng vị trí, ø(x) là hàm biến đổi và được gọi là Wavelet mẹ.

Phép biến đổi Wavelet rời rạc:

Để tính các hệ số của phép biến đổi Wavelet liên tục trên máy tính, hai tham số tỉ lệ và dịch chuyển không thể nhận các giá trị liên tục mà phải là các giá trị rời rạc. Năm 1989, Mallat đưa ra kỹ thuật phân tích đa phân giải trên cơ sở mã hóa hình tháp và đề xuất các họ hàm Wavelet trực giao để áp dụng trong xử lý tín hiệu số (Mallat, 1989). Ý tưởng của phân tích đa phân giải là sử dụng các kỹ thuật lọc trong quá trình phân tích. Trong đó, mỗi một tín hiệu được phân tích thành hai thành phần: thành phần xấp xỉ A tương ứng với thành phần tần số thấp và thành phần chi tiết D tương ứng với thành phần tần số cao.

Một tín hiệu x có độ dài *N*, phép biến đổi Wavelet rời rạc (DWT) bao gồm tối đa $\log_2 N$ tầng. Tầng đầu tiên, bắt đầu từ *x* tạo ra hai thành phần: thành phần xấp xỉ a_0 thu được bằng cách nhân chập x với bộ lọc thông thấp và thành phần chi tiết d_0 thu được bằng cách nhân chập x với bộ lọc thông cao. Trong đó, bộ lọc thông thấp sử dụng hàm tỉ lệ Φ(x) và bộ lọc thông cao sử dụng hàm Wavelet Ψ(x).

Mối quan hệ giữa hàm tỉ lệ và hàm Wavelet được cho bởi:

$$\Phi(x) = \sum_{k=0}^{N-1} C_k \Phi(2x-k) \quad (3)$$

$$\Psi(x) = \sum_{k=0}^{N-1} (-1)^k C_k \times \Phi(2x+k-N+1) \quad (4)$$

Trong đó: C là hằng số phụ thuộc vào hàm Wavelet được sử dụng.

Các phép lọc được tiến hành với nhiều tầng khác nhau và để khối lượng tính toán không tăng, khi qua mỗi bộ lọc, tín hiệu được lấy mẫu xuống 2. Tầng tiếp theo, lặp lại các bước như trên để chia thành phần xấp xỉ a_0 thành hai phần a_1 và d_1 và tiếp tục cho các tầng phía sau như trong hình 4.

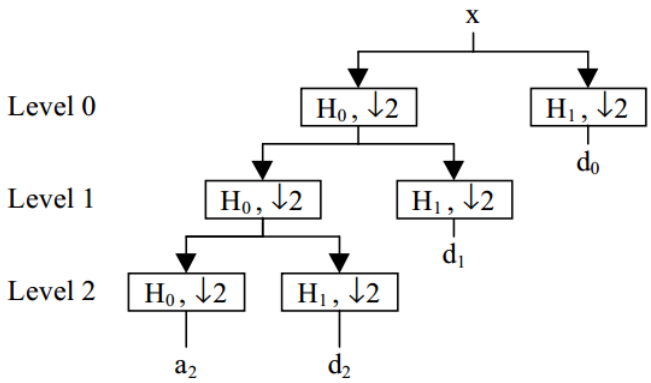
Tại mỗi tầng lọc, biểu thức của phép lọc được cho bởi công thức:

$$y_{\text{tần số cao}}(n) = \sum_n S(n)g(2k-n) \quad (5)$$

$$y_{\text{tần số thấp}}(n) = \sum_n S(n)h(2k-n) \quad (6)$$

Trong đó: S(n) là tín hiệu, h(n) là đáp ứng xung của các bộ lọc thông thấp tương ứng với hàm tỉ lệ Φ(n) và g(n) là đáp ứng xung của các bộ lọc thông cao tương ứng với hàm Wavelet Ψ(n). Hai bộ lọc này liên hệ nhau theo hệ thức:

$$h(N-n-1) = (-1)^n g(n) \quad (7)$$



Hình 4. Phân tích đa phân giải sử dụng biến đổi wavelet rời rạc

c. Chroma

Đặc trưng về cao độ là đặc trưng được sử dụng để đánh giá âm thanh là "cao" hay "thấp" theo nghĩa liên quan đến giai điệu âm nhạc. Nhận thức của con người về cao độ là tuần hoàn nghĩa là hai cao độ được coi là giống nhau về "màu sắc" nếu chúng khác nhau một hoặc vài quãng tám. Trong thang âm, một quãng tám được định nghĩa là khoảng cách của 12 nốt nhạc được biểu diễn bởi tập hợp {C, C♯, D, D♯, E ,F, F♯, G, G♯, A, A♯, B}. Cao độ được tách thành hai thành phần, được gọi là độ cao âm (tone height) và sắc độ (chroma). Độ cao âm đề cập đến số quãng tám và sắc độ là một bộ gồm 12 phần tử cho biết mức độ năng lượng của mỗi lớp cao độ {C, C♯, D,..., B} có trong tín hiệu (Kattel & cs., 2019).

Đặc trưng sắc độ được trích xuất từ cường độ quang phổ bằng cách sử dụng biến đổi Short Time Fourier Transform (STFT), Constant-Q transform(CQT),...

2.2.2. Một số phương pháp học máy

a. Máy véctơ hỗ trợ

Thuật toán máy véctơ hỗ trợ (SVM - Support Vector Machine), được Vapnik và các cộng sự của ông giới thiệu lần đầu tiên vào cuối thập kỷ 70 của thế kỷ XX, là một trong những thuật toán học tập dựa trên hàm nhân được sử dụng rộng rãi trong nhiều ứng dụng học máy. SVM thuộc nhóm các kỹ thuật học có giám sát phi tham số, không nhạy cảm với việc phân phối dữ liệu cơ bản. Đây là một trong những ưu điểm của SVM so với các kỹ thuật thống kê khác (Mather & Tso, 2016).

SVM là một bộ phân lớp nhị phân tuyến tính, xác định một ranh giới duy nhất giữa hai lớp. Để tối đa hóa sự phân tách, SVM sử dụng một phần của các mẫu huấn luyện nằm gần nhất với đường biên trong không gian đặc trưng như các véctơ hỗ trợ (support vector) (Kuo & cs., 2013). Ví dụ như trong hình 5, một siêu phẳng tối ưu được xác định với biên độ phân tách là cực đại.

Trong thực tế, các mẫu dữ liệu của các lớp khác nhau không phải lúc nào cũng phân tách tuyến tính mà có sự chồng chéo với nhau (Hình 6). Trong trường hợp này, chúng ta có thể tìm cách biến đổi dữ liệu sang một không gian mới sao cho ở đó dữ liệu của hai lớp là phân biệt tuyến tính hoặc gần phân biệt tuyến tính. Để thực hiện việc này, ta có thể sử dụng kernel trick. Có một số mô hình hạt nhân để xây dựng các SVM khác nhau, điển hình như: hàm Sigmoid, Hàm Radial basis (hay Gaussian), hàm tuyến tính, hàm đa thức.

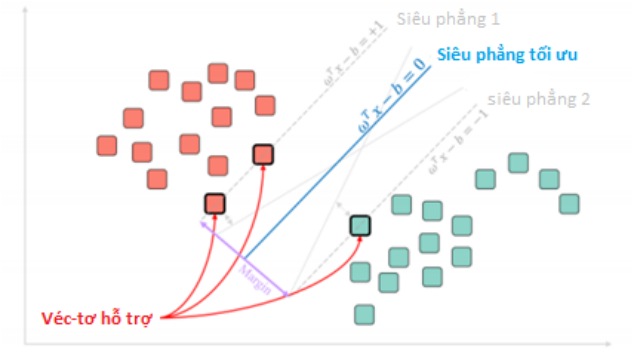
Để mở rộng mô hình SVM áp dụng cho bài toán phân lớp đa lớp, ta sử dụng nhiều bộ phân lớp nhị phân và các kỹ thuật như one-vs-one hoặc one-vs-all.

b. Phương pháp học kết hợp

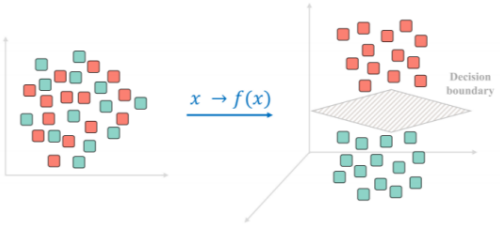
Học kết hợp là một thuật toán kết hợp một số thuật toán học máy vào một mô hình dự đoán để cải thiện kết quả dự đoán. Hai nhóm điển hình của phương pháp học kết hợp là Bagging và Boosting.

Rừng ngẫu nhiên

Rừng ngẫu nhiên (RF) là phương pháp học kết hợp có giám sát theo phương pháp Bagging, được phát triển bởi (Breiman, 2001) để có thể giải quyết cả hai bài toán phân lớp và hồi quy. Thuật toán RF bao gồm nhiều cây quyết định, mỗi cây được xây dựng dùng thuật toán cây quyết định trên tập dữ liệu khác nhau và dùng tập thuộc tính khác nhau. Sau đó kết quả dự đoán của thuật toán RF sẽ được tổng hợp từ các cây quyết định bằng cách lấy trung bình dự đoán của các cây quyết định trong trường hợp hồi quy hoặc sử dụng biểu quyết đa số cho nhãn lớp trong trường hợp phân lớp. Do mỗi cây quyết định đều có những yếu tố ngẫu nhiên và mỗi cây quyết không dùng tất cả dữ liệu huấn luyện, cũng như không dùng tất cả các đặc trưng của dữ liệu để xây dựng cây nên mỗi cây có thể sẽ dự đoán không tốt, khi đó mỗi mô hình cây quyết định không bị overfitting mà có thế bị underfitting. Tuy nhiên, kết quả cuối cùng của thuật toán RF lại tổng hợp từ nhiều cây quyết định, thế nên thông tin từ các cây sẽ bổ sung thông tin cho nhau, dẫn đến mô hình có kết quả dự đoán tốt (Vũ Hữu Tiệp, 2021).



Hình 5. Ví dụ SVM cho dữ liệu có thể phân tách tuyến tính



Hình 6. Ví dụ SVM cho dữ liệu không phân tách tuyến tính

Extremely Randomized Trees

Tương tự như RF, Extremely Randomized Trees (viết ngắn gọn là Extra Trees) là thuật toán học máy kết hợp hoạt động bằng cách tạo ra một số lượng lớn các cây quyết định. Tuy nhiên, hai điểm khác biệt chính của Extra Trees với RF là nó phân chia các nút bằng cách chọn các điểm cắt hoàn toàn ngẫu nhiên và nó sử dụng toàn bộ dữ liệu huấn luyện để phát triển cây (Geurts & cs., 2006).

XGBoost

XGBoost (Extreme Gradient Boosting) là một giải thuật được xây dựng dựa trên gradient boosting. Phương pháp này xây dựng một chuỗi các mô hình, trong đó mỗi mô hình sau cố gắng hạn chế lỗi của mô hình trước đó và do đó mô hình sau sẽ tốt hơn mô hình trước. Kết quả của mô hình cuối cùng trong chuỗi mô hình này sẽ là kết quả trả về. XGBoost có thể được sử dụng để giải quyết nhiều bài toán như hồi quy, phân lớp hoặc giải quyết các vấn đề do người dùng tự định nghĩa.

2.2.3. Thử nghiệm

Chúng tôi sử dụng hai tập dữ liệu BUZZ1 và BUZZ2 như đã mô tả ở trên để thử nghiệm. Trong đó, tập dữ liệu BUZZ2 đã được chia thành hai tập huấn luyện và thử nghiệm riêng biệt. Tập dữ liệu BUZZ1 gồm tất cả 9.110 mẫu, được chia thành tập huấn luyện gồm 6.103 mẫu (chiếm 67%) và tập thử nghiệm gồm 3.007 mẫu (chiếm 33%) với thủ tục *train_test_split* từ thư viện *sklearn.model_selection* của Python (Scikit-Learn, 2021).

Trước tiên, chúng tôi thử nghiệm với các phương pháp trích chọn đặc trưng cơ bản: MFCC, Wavelet và Chroma trên hai tập dữ liệu. Việc lựa chọn các phương pháp trích chọn này do chúng tôi nhận xét dữ liệu được sử dụng đã được thu thập từ tổ ong trong môi trường đô thị, trong đó có tiếng ồn là giọng nói người và MFCC được sử dụng rộng rãi trong nhận diện giọng nói và âm thanh vì khả năng mô phỏng cách tai người cảm nhận âm thanh; tiếng ong vo ve, tiếng dế kêu có thể giống giai điệu âm nhạc nên Wavelet và Chroma có thể sẽ hữu ích. Sau đó sử dụng các đặc trưng này để huấn luyện và thử nghiệm với các mô hình: SVM với nhân tuyến tính one vs. one, Rừng ngẫu nhiên và XGBoost. Các phương pháp được chúng tôi lựa chọn đại diện cho các phương pháp học đơn và hai phương pháp học kết hợp là Bagging và Boosting. Các

phương pháp này đều là mô hình phân lớp có giám sát được huấn luyện dựa trên dữ liệu được gắn nhãn để phân lớp dữ liệu quan sát mới vào một trong các lớp được xác định trước.

Tiếp theo, chúng tôi cài đặt thử nghiệm sử dụng các mô hình Rừng ngẫu nhiên, Extra Trees và XGBoost để lựa chọn ra các đặc trưng quan trọng của dữ liệu thô ban đầu và cũng sử dụng những đặc trưng này cho các mô hình học máy như trường hợp trên.

Chúng tôi đã sử dụng hai số liệu để đánh giá hiệu suất của các mô hình: độ chính xác của phân lớp và ma trận nhầm lẫn. Độ chính xác của phân lớp là tỷ lệ phần trăm các dự đoán đúng và ma trận nhầm lẫn chỉ ra cụ thể mỗi loại được phân loại như thế nào, lớp nào được phân loại đúng nhiều nhất và dữ liệu thuộc lớp nào thường bị phân loại nhầm vào lớp khác.

3. KẾT QUẢ VÀ THẢO LUẬN

Kết quả thực nghiệm với tập dữ liệu BUZZ1 được cho trong các bảng từ bảng 1 đến bảng 4:

Kết quả bảng 1 cho thấy với đặc trưng được trích chọn bởi kỹ thuật MFCC tất cả các phương pháp học máy kể trên đều cho kết quả tốt đặc biệt với hai mô hình Rừng ngẫu nhiên và XGBoost (≥ 99,9%). Đứng thứ 2 là kết quả phân lớp khi sử dụng các đặc trưng được trích chọn bởi phương pháp Chroma: kết quả phân lớp tốt nhất là phương pháp Rừng ngẫu nhiên 97,21%, tiếp theo là phương pháp XGBoost và cuối cùng là phương pháp SVM. Khi sử dụng đặc trưng Wavelet, mặc dù có số chiều rất lớn (88.200 chiều) nhưng hiệu quả phân lớp của các phương pháp học máy lại không tốt bằng sử dụng đầu vào là đặc trưng trích chọn bởi phương pháp MFCC và Chroma. Khả năng phân loại chính xác các dữ liệu đầu vào thuộc về phương pháp XGBoost 90,46%. Từ kết quả trên chúng ta thấy, MFCC là phương pháp trích chọn được các đặc trưng dữ liệu tốt hơn hai phương pháp trích chọn đặc trưng Chroma và Wavelet.

Ngoài việc so sánh độ chính xác phân lớp, chúng tôi còn thực hiện kiểm tra sự phân lớp nhầm lẫn giữa các loại dữ liệu khác nhau của các phương pháp phân lớp.

Số liệu trong bảng 2 cho thấy khi sử dụng đặc trưng MFCC và mô hình XGBoost để phân lớp cho 3.007 mẫu âm thanh thuộc tập dữ liệu BUZZ1 chỉ có 2 mẫu âm thanh thuộc lớp tiếng ong được phân loại nhầm sang lớp tiếng ồn, tất cả các mẫu còn lại đều được phân lớp đúng.

Bảng 1. Kết quả thử nghiệm trên tập dữ liệu BUZZ1 với các phương pháp trích chọn đặc trưng cơ bản

Phương pháp trích chọn đặc trưng	Số lượng đặc trưng	SVM (%)	Rừng ngẫu nhiên (%)	XGBoost (%)
MFCC	40	99,34	99,93	99,93
Chroma	12	96,41	97,21	97,17
Wavelet	88200	86,56	87,03	90,46

Bảng 2. Ma trận nhầm lẫn khi thử nghiệm với đặc trưng MFCC và phương pháp phân lớp XGBoost trên tập BUZZ1

	Kết quả phân lớp				
	Tiếng ong	Tiếng ồn	Tiếng dế	Tổng	Độ chính xác
Tiếng ong	1004	2	0	1006	99,80%
Tiếng ồn	0	1016	0	1016	100,00%
Tiếng dế	0	0	985	985	100,00%
Độ chính xác					99,93%

Bảng 3. Kết quả thử nghiệm trên tập dữ liệu BUZZ1 với các phương pháp học máy được sử dụng để lựa chọn ra các đặc trưng quan trọng

Phương pháp lựa chọn đặc trưng	Số lượng đặc trưng	SVM (%)	Rừng ngẫu nhiên (%)	XGBoost (%)
Rừng ngẫu nhiên	87488	89,26	86,66	91,22
Extra trees	2649	88,40	87,20	88,53
XGBoost	7039	89,66	87,40	91,02

Bảng 4. Ma trận nhầm lẫn khi sử dụng XGBoost để lựa chọn đặc trưng và phương pháp phân lớp Rừng ngẫu nhiên trên tập BUZZ1

	Kết quả phân lớp				
	Tiếng ong	Tiếng ồn	Tiếng dế	Tổng	Độ chính xác
Tiếng ong	763	149	76	988	77,23%
Tiếng ồn	17	1001	10	1.028	97,37%
Tiếng dế	127	0	864	991	87,18%
Độ chính xác					87,40%

Bảng 5. Kết quả thử nghiệm trên tập dữ liệu BUZZ2 với các phương pháp trích chọn đặc trưng cơ bản

Phương pháp trích chọn đặc trưng	Số lượng đặc trưng	SVM (%)	Rừng ngẫu nhiên (%)	XGBoost (%)
MFCC	40	99,44	98,85	99,19
Chroma	12	88,34	87,95	89,15
Wavelet	88.200	40,05	39,97	39,45

Bảng 3 là kết quả so sánh của các phương pháp học máy khi sử dụng các phương pháp Rừng ngẫu nhiên, Extra trees, XGBoost để lựa chọn các đặc trưng quan trọng trong số 88.200 đặc trưng thô ban đầu. Kết quả trong bảng 3 cho thấy phương pháp XGBoost luôn cho kết quả tốt nhất, đặc biệt với đặc trưng được trích chọn sử dụng Rừng ngẫu nhiên (≥ 99,2%). Tuy nhiên, kết quả vẫn không tốt bằng khi sử dụng đặc trưng MFCC (≥ 99,9%) mặc dù đặc trưng MFCC có kích thước là 40 nhỏ hơn nhiều so với các phương pháp trích chọn đặc trưng này

Bảng 4 thể hiện phân lớp nhầm lẫn giữa các loại dữ liệu khác nhau khi sử dụng XGBoost để lựa chọn đặc trưng và phương pháp phân lớp Rừng ngẫu nhiên. Với tổng số 3.007 mẫu đã có: 149 mẫu âm thanh tiếng ong bị phân nhầm vào lớp tiếng ồn, 76 mẫu âm thanh tiếng ong bị phân nhầm vào lớp tiếng dế; 17 mẫu thuộc lớp tiếng ồn bị phân vào lớp tiếng ong, 10 mẫu tiếng ồn bị phân vào lớp tiếng dế và 127 mẫu tiếng dế lại được phân vào lớp tiếng ong.

Kết quả thực nghiệm với tập dữ liệu BUZZ2 được cho trong các bảng từ bảng 5 đến bảng 8.

Kết quả trong bảng 5 cho thấy phương pháp trích chọn đặc trưng MFCC cho kết quả phân lớp tốt nhất (99,44%) khi sử dụng bộ phân lớp SVM; phương pháp trích chọn đặc trưng Chroma cho kết quả phân lớp tốt nhất (89,15%) khi sử dụng bộ phân lớp XGBoost và phương pháp trích chọn đặc trưng Wavelet cho kết quả phân lớp thấp hơn nhiều so với hai phương trên, kết quả tốt nhất chỉ đạt 40,05% khi sử dụng bộ phân lớp SVM. So sánh kết quả trên tập BUZZ2 với tập BUZZ1 ta thấy hai phương pháp MFCC và Chroma cho kết quả không thấp hơn nhiều; tuy nhiên phương pháp Wavelet cho kết quả tốt trên tập BUZZ1 (90,46%) nhưng lại cho kết quả thấp trên tập BUZZ2 (40,05%).

Bảng 6. Ma trận nhầm lẫn khi thử nghiệm với đặc trưng MFCC và phương pháp phân lớp XGBoost trên tập BUZZ2

	Kết quả phân lớp				
	Tiếng ong	Tiếng ồn	Tiếng dế	Tổng	Độ chính xác
Tiếng ong	893	5	0	898	99,44%
Tiếng ồn	1	930	3	934	99,57%
Tiếng dế	0	10	490	500	98,00%
Độ chính xác					99,19%

Bảng 7. Kết quả thử nghiệm trên tập dữ liệu BUZZ2 với các phương pháp học máy được sử dụng để lựa chọn ra các đặc trưng quan trọng

Phương pháp lựa chọn đặc trưng	Số lượng đặc trưng	SVM (%)	Rừng ngẫu nhiên (%)	XGBoost (%)
Rừng ngẫu nhiên	26683	40.05	40.00	39.49
Extra trees	2649	40.05	39.97	39.45
XGBoost	4891	40.05	39.97	39.40

Bảng 8. Ma trận nhầm lẫn khi sử dụng XGBoost để lựa chọn đặc trưng và phương pháp phân lớp Rừng ngẫu nhiên trên tập BUZZ2

	Kết quả phân lớp				
	Tiếng ong	Tiếng ồn	Tiếng dế	Tổng	Độ chính xác
Tiếng ong	0	898	0	898	0,00%
Tiếng ồn	2	932	0	934	99,79%
Tiếng dế	0	500	0	500	0,00%
Độ chính xác					39,97%

Số liệu trong bảng 6 cho thấy với đặc trưng MFCC và bộ phân lớp XGBoost trong số 2.332 mẫu âm thanh thuộc bộ BUZZ2 chỉ có 19 mẫu bị phân lớp sai.

Số liệu trong bảng 7 cho thấy, khi sử dụng các phương pháp Rừng ngẫu nhiên, Extra trees, XGBoost để lựa chọn các đặc trưng và kết quả phân lớp thu được đều rất thấp với cả ba phương pháp phân lớp. Kết quả này cũng thấp hơn nhiều khi ta so sánh với kết quả trên tập BUZZ1 như thể hiện trong bảng 3.

Kết quả thể hiện trong bảng 8 cho thấy khi sử dụng XGBoost để lựa chọn đặc trưng và phương pháp phân lớp Rừng ngẫu nhiên trên tập BUZZ2, tất cả các mẫu âm thanh thuộc lớp tiếng ong và lớp tiếng dế đều bị xếp vào lớp tiếng ồn.

4. KẾT LUẬN

Từ kết quả thử nghiệm trên, chúng tôi nhận xét rằng hai phương pháp trích chọn đặc trưng MFCC và Chroma cho kết quả phân lớp tốt trên cả ba phương pháp học máy SVM, Rừng ngẫu nhiên và XGBoost và trên cả hai tập dữ liệu khác nhau. Đặc trưng MFCC cho kết quả tốt nhất trong số các phương pháp trích chọn đặc trưng đã được thử nghiệm, như vậy chúng ta hoàn toàn có thể sử dụng đặc trưng này để ứng dụng vào bài toán giám sát tổ ong dựa trên âm thanh ong trong thực tế.

Với các phương pháp còn lại: Wavelet, Rừng ngẫu nhiên, Extra trees, XGBoost mặc dù cho kết quả phân lớp tốt trên tập BUZZ1 nhưng lại cho kết quả không tốt trên tập dữ liệu BUZZ2,

là tập dữ liệu mà hai tập huấn luyện và kiểm thử hoàn toàn tách biệt nhau về tổ ong và vị trí đặt tổ. Nhóm sẽ thực hiện các nghiên cứu để cải tiến các phương pháp trích chọn đặc trưng này trong các nghiên cứu tiếp theo.

Chúng tôi đã thử nghiệm với nhiều kiểu đặc trưng khác nhau và kết quả thử nghiệm cho thấy: trên tập dữ liệu BUZZ1, với hai mô hình Rừng ngẫu nhiên và XGBoost khi sử dụng đặc trưng MFCC cho kết quả phân lớp chính xác tới 99,93% tương đương kết quả cho trong bảng 3 khi sử dụng CNN; trên tập dữ liệu BUZZ2, kết quả phân lớp khi sử dụng đặc trưng MFCC và SVM là 99,44% tốt hơn kết quả cho trong bảng 14 (~ 95,7%) của nhóm tác giả Kulyukin & cs. (2018). Từ đó chúng tôi rút ra nhận xét, với dữ liệu này chỉ cần sử dụng đặc trưng MFCC và các kỹ thuật truyền thống là đã đáp ứng được yêu cầu của bài toán, mà không cần tốn nhiều chi phí tính toán như dùng CNN.

LỜI CẢM ƠN

Nghiên cứu này được hỗ trợ bởi đề tài "Nghiên cứu ứng dụng công nghệ của công nghiệp 4.0 vào quản lý sản xuất sản phẩm mật ong phục vụ xuất khẩu và tiêu dùng trong nước", mã số KC4.0-20/19-25."

TÀI LIỆU THAM KHẢO

Abdalla M.I. & Ali H.S. (2010). Wavelet-Based Mel-Frequency Cepstral Coefficients for Speaker Identification using Hidden Markov Models. Telecommunications. 1(2): 16-21.

Breiman L. (2001). Random forests. Machine Learning. 45(1): 5-32.

Bromenshenk J.J., Henderson C.B., Seccomb R.A., Rice S.D. & Etter R.T. (2009). Honey bee acoustic recording and analysis system for monitoring hive health. Google Patents.

Davis S. & Mermelstein P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences, IEEE Transactions on Acoustics, Speech, and Signal Processing. 28(4): 357-366.

Du N.H., Dong N.D., Luu V.T., Hoang N.V., Thai P.H., Ngoc P.T., Long N.V. & Hong P.T.T. (2020). Toward Audio Beehive Monitoring Based on IoT-AI techniques: A Survey and Perspective, Vietnam Journal of Agricultural Sciences. 3(1): 530-540.

Geurts P., Ernst D. & Wehenkel L. (2006). Extremely randomized trees. Machine learning, 63(1): 3-42.

Kattel M., Nepal A., Shah A.K. & Shrestha D. (2019). Chroma feature extraction. Conference: Chroma Feature Extraction using Fourier Transform. 20(1).

Kulyukin V. (2018). BeePi: A Multisensor Electronic Beehive Monitor Retrieved. Truy cập từ https://www.kickstarter.com/projects/beepihoneybeesmeetai/beepi-a-multisensor-electronic-beehive-monitor ngày 10/9/2021.

Kulyukin V., Mukherjee S. & Amlathe P. (2018). Toward audio beehive monitoring: Deep learning vs. standard machine learning in classifying beehive audio samples. Applied Sciences. 8(9): 1573.

Kuo B.C., Ho H.H., Li C.H., Hung C.C. & Taur J.S. (2013). A kernel-based feature selection method for SVM with RBF kernel for hyperspectral image classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 7(1): 317-326.

Mallat S.G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. IEEE Transactions on Pattern Analysis and Machine Intelligence. 11(7): 674-693.

Mather P. & Tso B. (2016). Classification methods for remotely sensed data. CRC press.

Nolasco I., Terenzi A., Cecchi S., Orcioni S., Bear H.L. & Benetos E., (2019). Audio-based identification of beehive states. ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE. pp. 8256-8260.

Python (2021). Truy cập từ https://www.python.org/ ngày 20/6/2021.

Robles-Guerrero A., Saucedo-Anaya T., González-Ramérez E. & Galván-Tejada C.E. (2017). Frequency Analysis of Honey Bee Buzz for Automatic Recognition of Health Status: A Preliminary Study. Research in Computing Science 142: 89-98.

Scikit-Learn (2021). Truy cập từ https://scikit-learn.org ngày 20/6/2021.

Terenzi A., Cecchi S., Orcioni S. & Piazza F. (2019). Features extraction applied to the analysis of the sounds emitted by honey bees in a beehive. 2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA). IEEE. pp. 03-08.

Vũ Hữu Tiệp (2021). Random Forest algorithm. Truy cập từ https://machinelearningcoban.com/ tabml _book/ch_model/random_forest.html ngày 12/5/2021.